Problem 2: Off Policy Evaluation and Causal Inference

(a) Importance Sampling

Given: The importance sampling estimator

$$\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\left[\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}R(s,a)\right]$$

We need to show that if $\hat{\pi}_0 = \pi_0$, then this equals $\mathbb{E}_{s\sim p(s),a\sim\pi_1(s,a)}[R(s,a)]$.
Proof:

$$\begin{aligned}
&\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\left[\frac{\pi_1(s,a)}{\pi_0(s,a)}R(s,a)\right] \\
&= \sum_{s,a} p(s)\pi_0(s,a)\cdot\frac{\pi_1(s,a)}{\pi_0(s,a)}R(s,a) \\
&= \sum_{s,a} p(s)\pi_1(s,a)R(s,a) \\
&= \mathbb{E}_{s\sim p(s),a\sim\pi_1(s,a)}[R(s,a)]
\end{aligned}$$

(b) Weighted Importance Sampling

Given: The weighted importance sampling estimator

$$\frac{\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\left[\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}R(s,a)\right]}{\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\left[\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}\right]}$$

We need to show that if $\hat{\pi}_0 = \pi_0$, then this equals $\mathbb{E}_{s\sim p(s),a\sim\pi_1(s,a)}[R(s,a)]$.
Proof:
Numerator:

$$\begin{aligned}
\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\left[\frac{\pi_1(s,a)}{\pi_0(s,a)}R(s,a)\right] &= \sum_{s,a} p(s)\pi_0(s,a)\cdot\frac{\pi_1(s,a)}{\pi_0(s,a)}R(s,a) \\
&= \sum_{s,a} p(s)\pi_1(s,a)R(s,a)
\end{aligned}$$

Denominator:

$$\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\left[\frac{\pi_1(s,a)}{\pi_0(s,a)}\right] = \sum_{s,a} p(s)\pi_0(s,a)\cdot\frac{\pi_1(s,a)}{\pi_0(s,a)}$$
$$= \sum_{s,a} p(s)\pi_1(s,a)$$
$$= \sum_{s} p(s)\sum_{a}\pi_1(s,a)$$
$$= \sum_{s} p(s)\cdot 1 = 1$$

Therefore:

$$\frac{\sum_{s,a} p(s)\pi_1(s,a)R(s,a)}{1} = \mathbb{E}_{s\sim p(s),a\sim\pi_1(s,a)}[R(s,a)]$$

(c) Bias in Weighted Importance Sampling

Consider a dataset with a single observation $(s_0, a_0, R(s_0, a_0))$.

The weighted importance sampling estimator becomes:

$$\frac{\frac{\pi_1(s_0,a_0)}{\pi_0(s_0,a_0)}R(s_0,a_0)}{\frac{\pi_1(s_0,a_0)}{\pi_0(s_0,a_0)}} = R(s_0,a_0)$$

However, the true expected value is:

$$\mathbb{E}_{s\sim p(s),a\sim\pi_1(s,a)}[R(s,a)] = \sum_{s,a} p(s)\pi_1(s,a)R(s,a)$$

Since $R(s_0, a_0)$ is just one sample and may not equal the full expectation over all possible states and actions, the estimator is biased.

(d) Doubly Robust Estimator

Given:

$$\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\left[\mathbb{E}_{a\sim\pi_1(s,a)}[\hat{R}(s,a)] + \frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}(R(s,a) - \hat{R}(s,a))\right]$$

(i) When $\hat{\pi}_0 = \pi_0$

$$\begin{aligned}
&\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\left[\mathbb{E}_{a\sim\pi_1(s,a)}[\hat{R}(s,a)] + \frac{\pi_1(s,a)}{\pi_0(s,a)}(R(s,a) - \hat{R}(s,a))\right] \\
&= \mathbb{E}_{s\sim p(s)}\left[\mathbb{E}_{a\sim\pi_1(s,a)}[\hat{R}(s,a)]\right] + \mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\left[\frac{\pi_1(s,a)}{\pi_0(s,a)}(R(s,a) - \hat{R}(s,a))\right]
\end{aligned}$$

For the second term:

$$\begin{aligned}
&\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\left[\frac{\pi_1(s,a)}{\pi_0(s,a)}R(s,a)\right] - \mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\left[\frac{\pi_1(s,a)}{\pi_0(s,a)}\hat{R}(s,a)\right] \\
&= \mathbb{E}_{s\sim p(s),a\sim\pi_1(s,a)}[R(s,a)] - \mathbb{E}_{s\sim p(s),a\sim\pi_1(s,a)}[\hat{R}(s,a)]
\end{aligned}$$

Combining:

$$\begin{aligned}
&\mathbb{E}_{s\sim p(s),a\sim\pi_1(s,a)}[\hat{R}(s,a)] + \mathbb{E}_{s\sim p(s),a\sim\pi_1(s,a)}[R(s,a)] - \mathbb{E}_{s\sim p(s),a\sim\pi_1(s,a)}[\hat{R}(s,a)] \\
&= \mathbb{E}_{s\sim p(s),a\sim\pi_1(s,a)}[R(s,a)]
\end{aligned}$$

(ii) When $\hat{R}(s,a) = R(s,a)$

When $\hat{R}(s,a) = R(s,a)$, the second term becomes zero:

$$\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}(R(s,a) - \hat{R}(s,a)) = 0$$

Therefore:

$$\begin{aligned}
&\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\left[\mathbb{E}_{a\sim\pi_1(s,a)}[R(s,a)]\right] \\
&= \mathbb{E}_{s\sim p(s)}\left[\mathbb{E}_{a\sim\pi_1(s,a)}[R(s,a)]\right] \\
&= \mathbb{E}_{s\sim p(s),a\sim\pi_1(s,a)}[R(s,a)]
\end{aligned}$$

(e) Comparison of Estimators

(i) Random drug assignment, complicated interaction

Importance Sampling would work better.

Since drugs are randomly assigned, π_0 is simple and easy to estimate accurately. However, the complicated interaction makes it difficult to model $R(s,a)$ accurately. Therefore, importance sampling (which only requires modeling π_0) is preferred over regression (which requires modeling $R(s,a)$).

(ii) Complicated drug assignment, simple interaction

Regression would work better.

When the assignment policy π_0 is complicated, it's difficult to estimate accurately. However, if the interaction between drug, patient, and lifespan is simple, we can easily model $R(s, a)$ accurately. Therefore, regression estimator is preferred.

Problem 3: PCA - Variance Maximizing Interpretation

We need to show that:

$$\arg\min_{u:u^T u=1} \sum_{i=1}^{m} \|x^{(i)} - f_u(x^{(i)})\|_2^2$$

gives the first principal component.

Solution

$$f_u(x) = (u^T x)u$$

$$\begin{aligned}\sum_{i=1}^{m} \|x^{(i)} - f_u(x^{(i)})\|_2^2 &= \sum_{i=1}^{m} \|x^{(i)} - (u^T x^{(i)})u\|_2^2 \\ &= \sum_{i=1}^{m} (x^{(i)} - (u^T x^{(i)})u)^T (x^{(i)} - (u^T x^{(i)})u)\end{aligned}$$

$$= \sum_{i=1}^{m} \left[(x^{(i)})^T x^{(i)} - (x^{(i)})^T (u^T x^{(i)})u - ((u^T x^{(i)})u)^T x^{(i)} + ((u^T x^{(i)})u)^T ((u^T x^{(i)})u)\right]$$

$(u^T x^{(i)})$ is a scalar, so:

$$\begin{aligned}(x^{(i)})^T (u^T x^{(i)})u &= (u^T x^{(i)})(x^{(i)})^T u = (u^T x^{(i)})^2 \\ ((u^T x^{(i)})u)^T x^{(i)} &= (u^T x^{(i)})u^T x^{(i)} = (u^T x^{(i)})^2 \\ ((u^T x^{(i)})u)^T ((u^T x^{(i)})u) &= (u^T x^{(i)})^2 u^T u = (u^T x^{(i)})^2\end{aligned}$$

where we used $u^T u = 1$ in the last step.

$$
\begin{aligned}
&= \sum_{i=1}^{m} \left[(x^{(i)})^T x^{(i)} - 2(u^T x^{(i)})^2 + (u^T x^{(i)})^2 \right] \\
&= \sum_{i=1}^{m} \left[(x^{(i)})^T x^{(i)} - (u^T x^{(i)})^2 \right] \\
&= \sum_{i=1}^{m} \|x^{(i)}\|_2^2 - \sum_{i=1}^{m} (u^T x^{(i)})^2
\end{aligned}
$$

Since $\sum_{i=1}^{m} \|x^{(i)}\|_2^2$ is constant (independent of u), minimizing the objective is equivalent to:

$$\arg\min_{u:u^Tu=1} \left(\sum_{i=1}^{m} \|x^{(i)}\|_2^2 - \sum_{i=1}^{m} (u^T x^{(i)})^2 \right)$$

This is the same as:

$$\arg\max_{u:u^Tu=1} \sum_{i=1}^{m} (u^T x^{(i)})^2$$

Let $X \in \mathbb{R}^{m \times d}$ be the design matrix where the i-th row is $(x^{(i)})^T$.

$$
\begin{aligned}
\sum_{i=1}^{m} (u^T x^{(i)})^2 &= \sum_{i=1}^{m} (x^{(i)})^T u u^T x^{(i)} \\
&= u^T \left(\sum_{i=1}^{m} x^{(i)} (x^{(i)})^T \right) u \\
&= u^T X^T X u
\end{aligned}
$$

Since the data is preprocessed to have zero mean, $\frac{1}{m} X^T X = \Sigma$ is the empirical covariance matrix.

Therefore:

$$\arg\max_{u:u^Tu=1} \sum_{i=1}^{m} (u^T x^{(i)})^2 = \arg\max_{u:u^Tu=1} u^T X^T X u = \arg\max_{u:u^Tu=1} u^T \Sigma u$$

The unit vector u that maximizes $u^T \Sigma u$ subject to $u^T u = 1$ is the eigenvector corresponding to the largest eigenvalue of Σ, which is exactly the first principal component.

Therefore, minimizing the mean squared error between projected points and original points gives us the first principal component of the data.

Problem 4: Independent Components Analysis

(a) Gaussian Source

Given that sources are distributed according to standard normal distribution: $s_j \sim \mathcal{N}(0,1)$ for $j = 1, \ldots, d$.

The likelihood function is:

$$\ell(W) = \sum_{i=1}^{n} \left[\log |W| + \sum_{j=1}^{d} \log g'(w_j^T x^{(i)}) \right]$$

For Gaussian distribution, the PDF is:

$$g'(s) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{s^2}{2}\right)$$

Therefore:

$$\log g'(w_j^T x^{(i)}) = \log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}(w_j^T x^{(i)})^2$$

$$\begin{aligned}
\ell(W) &= \sum_{i=1}^{n} \left[\log |W| + \sum_{j=1}^{d} \left(\log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}(w_j^T x^{(i)})^2 \right) \right] \\
&= n \log |W| + nd \log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{d} (w_j^T x^{(i)})^2
\end{aligned}$$

Note that:

$$\sum_{j=1}^{d} (w_j^T x^{(i)})^2 = \|W x^{(i)}\|_2^2 = (W x^{(i)})^T (W x^{(i)}) = (x^{(i)})^T W^T W x^{(i)}$$

Therefore:

$$\begin{aligned}
\ell(W) &= n \log |W| + nd \log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2} \sum_{i=1}^{n} (x^{(i)})^T W^T W x^{(i)} \\
&= n \log |W| + \text{const} - \frac{1}{2} \sum_{i=1}^{n} (x^{(i)})^T W^T W x^{(i)}
\end{aligned}$$

In matrix form, where $X \in \mathbb{R}^{n\times d}$ is the design matrix:

$$\sum_{i=1}^{n}(x^{(i)})^T W^T W x^{(i)} = \mathrm{tr}(XW^T W X^T) = \mathrm{tr}(WX^T X W^T)$$

To maximize $\ell(W)$, we need to minimize $\mathrm{tr}(WX^TXW^T)$ subject to maximizing $\log|W|$. Taking the gradient and setting to zero leads to the condition that W should satisfy:

$$W^T W = m(X^T X)^{-1}$$

The solution has **rotational invariance**. Any orthogonal rotation of W will give the same likelihood. This is because Gaussian distributions are rotationally symmetric, so we cannot uniquely determine W - any rotation of the unmixing matrix will produce sources that are also Gaussian and equally valid.

(b) Laplace Source

Given: $s_i \sim \mathcal{L}(0,1)$ with PDF $f_L(s) = \frac{1}{2}\exp(-|s|)$.

The likelihood function is:

$$\ell(W) = \sum_{i=1}^{n}\left[\log|W| + \sum_{j=1}^{d}\log g'(w_j^T x^{(i)})\right]$$

For Laplace distribution:

$$g'(s) = \frac{1}{2}e^{-|s|}$$

Therefore:

$$\log g'(w_j^T x^{(i)}) = \log\left(\frac{1}{2}\right) - |w_j^T x^{(i)}|$$

The likelihood becomes:

$$\ell(W) = n\log|W| + nd\log\left(\frac{1}{2}\right) - \sum_{i=1}^{n}\sum_{j=1}^{d}|w_j^T x^{(i)}|$$

For gradient ascent, we need:

$$\nabla_W \ell(W) = \nabla_W\left[n\log|W| - \sum_{i=1}^{n}\sum_{j=1}^{d}|w_j^T x^{(i)}|\right]$$

We know that:

$$\nabla_W \log |W| = (W^{-1})^T = (W^T)^{-1}$$

For the second term, note that:

$$\frac{\partial}{\partial w_j} |w_j^T x^{(i)}| = \text{sign}(w_j^T x^{(i)}) \cdot x^{(i)}$$

where $\text{sign}(z) = \begin{cases} 1 & \text{if } z > 0 \\ -1 & \text{if } z < 0 \end{cases}$

For a single training example $x^{(i)}$, the gradient is:

$$\nabla_W \ell^{(i)}(W) = (W^T)^{-1} - \begin{bmatrix} \text{sign}(w_1^T x^{(i)}) \cdot x^{(i)} \\ \text{sign}(w_2^T x^{(i)}) \cdot x^{(i)} \\ \vdots \\ \text{sign}(w_d^T x^{(i)}) \cdot x^{(i)} \end{bmatrix}^T$$

In more compact notation:

$$\nabla_W \ell^{(i)}(W) = (W^T)^{-1} - \text{sign}(W x^{(i)})(x^{(i)})^T$$

where sign is applied element-wise.

The stochastic gradient ascent update rule for a single example is:

$$\boxed{W := W + \alpha \left[(W^T)^{-1} - \text{sign}(W x^{(i)})(x^{(i)})^T \right]}$$

Alternatively, this can be written as:

$$\boxed{W := W + \alpha \left[(W^T)^{-1} - \begin{bmatrix} \text{sign}(w_1^T x^{(i)}) \\ \vdots \\ \text{sign}(w_d^T x^{(i)}) \end{bmatrix} (x^{(i)})^T \right]}$$

Problem 5: Markov Decision Processes

(a) Proving the Bellman operator is a γ-contraction

We need to prove that for any two finite-valued vectors V_1, V_2:

$$\|B(V_1) - B(V_2)\|_\infty \leq \gamma \|V_1 - V_2\|_\infty$$

where $\|V\|_\infty = \max_{s \in S} |V(s)|$.

Proof

Recall that the Bellman update operator is defined as:

$$B(V)(s) = R(s) + \gamma \max_{a\in A} \sum_{s'\in S} P_{sa}(s')V(s')$$

Let's compute $B(V_1)(s) - B(V_2)(s)$ for an arbitrary state s:

$$\begin{aligned}
B(V_1)(s) - B(V_2)(s) &= R(s) + \gamma \max_{a\in A} \sum_{s'\in S} P_{sa}(s')V_1(s') \\
&\quad - R(s) - \gamma \max_{a\in A} \sum_{s'\in S} P_{sa}(s')V_2(s') \\
&= \gamma \left[\max_{a\in A} \sum_{s'\in S} P_{sa}(s')V_1(s') - \max_{a\in A} \sum_{s'\in S} P_{sa}(s')V_2(s') \right]
\end{aligned}$$

We use the property that for any real numbers x_i and y_i:

$$\max_i x_i - \max_i y_i \le \max_i (x_i - y_i)$$

Similarly:

$$\max_i x_i - \max_i y_i \ge \min_i (x_i - y_i) \ge -\max_i |x_i - y_i|$$

Therefore:

$$\left| \max_i x_i - \max_i y_i \right| \le \max_i |x_i - y_i|$$

Applying this property:

$$\begin{aligned}
|B(V_1)(s) - B(V_2)(s)| &= \gamma \left| \max_{a\in A} \sum_{s'\in S} P_{sa}(s')V_1(s') - \max_{a\in A} \sum_{s'\in S} P_{sa}(s')V_2(s') \right| \\
&\le \gamma \max_{a\in A} \left| \sum_{s'\in S} P_{sa}(s')V_1(s') - \sum_{s'\in S} P_{sa}(s')V_2(s') \right| \\
&= \gamma \max_{a\in A} \left| \sum_{s'\in S} P_{sa}(s')(V_1(s') - V_2(s')) \right|
\end{aligned}$$

Using the triangle inequality and the fact that $\sum_{s'} P_{sa}(s') = 1$:

$$\begin{aligned}
\left|\sum_{s'\in S} P_{sa}(s')(V_1(s') - V_2(s'))\right| &\le \sum_{s'\in S} P_{sa}(s')|V_1(s') - V_2(s')| \\
&\le \sum_{s'\in S} P_{sa}(s')\|V_1 - V_2\|_\infty \\
&= \|V_1 - V_2\|_\infty \sum_{s'\in S} P_{sa}(s') \\
&= \|V_1 - V_2\|_\infty
\end{aligned}$$

Therefore:

$$|B(V_1)(s) - B(V_2)(s)| \le \gamma \max_{a\in A} \|V_1 - V_2\|_\infty = \gamma\|V_1 - V_2\|_\infty$$

Since this holds for all states $s \in S$:

$$\begin{aligned}
\|B(V_1) - B(V_2)\|_\infty &= \max_{s\in S} |B(V_1)(s) - B(V_2)(s)| \\
&\le \gamma\|V_1 - V_2\|_\infty
\end{aligned}$$

This completes the proof. $\square$

(b) Uniqueness of fixed point

We need to prove that B has at most one fixed point, i.e., there is at most one solution to the Bellman equations.

Proof by contradiction

Assume that there exist two distinct fixed points V^* and V^{**} such that:

$$B(V^*) = V^* \quad \text{and} \quad B(V^{**}) = V^{**}$$

and $V^* \neq V^{**}$ (i.e., $\|V^* - V^{**}\|_\infty > 0$).
Since both are fixed points, we have:

$$V^* - V^{**} = B(V^*) - B(V^{**})$$

Taking the infinity norm of both sides:

$$\|V^* - V^{**}\|_\infty = \|B(V^*) - B(V^{**})\|_\infty$$

From part (a), we know that B is a γ-contraction:

$$\|B(V^*) - B(V^{**})\|_\infty \leq \gamma \|V^* - V^{**}\|_\infty$$

Combining these results:

$$\|V^* - V^{**}\|_\infty \leq \gamma \|V^* - V^{**}\|_\infty$$

Since $\gamma < 1$ and $\|V^* - V^{**}\|_\infty > 0$, we can divide both sides by $\|V^* - V^{**}\|_\infty$:

$$1 \leq \gamma$$

Conclusion

This is a contradiction since $\gamma < 1$. Therefore, our assumption that there exist two distinct fixed points must be false.

Hence, B has **at most one fixed point**. Combined with the assumption that B has at least one fixed point, we conclude that B has **exactly one fixed point**.

Remark

This result guarantees that:

- The Bellman equations have a unique solution
- Value iteration converges to the unique optimal value function V^*
- The convergence is geometric with rate γ^k after k iterations